Assignment 02
PUBH 8878

Due before class on **Wednesday, September 10th**.

Requirements:

- Show all mathematical work
- Submit well-documented R code with clear comments
- Interpret results in biological context
- Submit *pdf* rendered output, not the source `.qmd` file

Problem 1: Missing Heritability and Rare Variants (40 pts)

Recall (Young 2019): "The first challenge is one of precision. The information used to estimate heritability from rare variants by GREML-WGS comes from the variation in sharing of rare variants among distantly related pairs of individuals. However, distantly related individuals typically do not share any particular rare variant, so the variation in rare variant sharing is low. This means that large samples with high quality WGS data are required to obtain precise estimates, and such samples are not common yet. Based on the only existing application of GREML-WGS, a sample size of ~40,000 would produce estimates precise enough to be statistically distinguished from other heritability estimates. It is likely that this challenge will be overcome shortly, since samples of similar magnitude already exist."

1. Assume the probability that two distantly related individuals share a rare variant is $p = 0.001$. Assume a sample size of $n = 40,000$ individuals.

    a. Calculate the expected number of pairs of individuals in this sample who share a rare variant. How many total pairs of individuals exist in the sample?
    b. If we estimate that rare variants contribute $h^2_{\text{rare}} = 0.10$ to heritability, calculate the standard error of this estimate given the sample size. Use the formula $\text{SE}(h^2) \approx \frac{2}{\sqrt{n_{\text{eff}}}}$, where n_{eff} is the effective number of independent observations (approximately the number of pairs sharing rare variants).
    c. Calculate a 95% confidence interval for the heritability estimate. Does this confidence interval allow us to distinguish between $h^2_{\text{rare}} = 0.10$ and $h^2_{\text{common}} = 0.25$

2. Briefly explain in 2-3 sentences how the "missing heritability" problem relates to rare variants, and why larger samples with whole-genome sequencing may be needed to resolve this question.

Problem 2: Parent–offspring regression with assortative mating (40 pts)

Let $Y = A + E$ with $\mathrm{Var}(A) = \sigma_A^2$, $\mathrm{Var}(E) = \sigma_E^2$, random environments, and phenotypic mate correlation $\mathrm{corr}(Y_{\text{father}}, Y_{\text{mother}}) = r_m$.

1. Show that $\mathrm{Var}(Y_{\text{mid-parent}}) = \frac{1}{2}(1 + r_m)\,\mathrm{Var}(Y)$.
2. Given the regression slope $\beta = \frac{\mathrm{Cov}(Y_o, Y_{mp})}{\mathrm{Var}(Y_{mp})}$, show that $\beta = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2} \cdot \frac{1}{1 + r_m} = \frac{h^2}{1 + r_m}$. Interpret the direction of bias in β for $r_m > 0$.
3. For $h^2 = 0.5$, compute β for $r_m = 0, 0.1, 0.3, 0.5$. Comment on the practical impact of assortative mating on parent–offspring regression.

Problem 3: Derivation of Falconer's Formula from the ACE Model (20 pts)

Background: The ACE model is a foundational tool in quantitative genetics for partitioning phenotypic variance (V_P) into three components: additive genetic effects (A), common or shared environmental effects (C), and unique or non-shared environmental effects (E). Under this model, the total variance is given by $V_P = V_A + V_C + V_E$.

The intraclass correlations for a trait between monozygotic (MZ) and dizygotic (DZ) twins are given by:

$$r_{\text{MZ}} = \frac{\mathrm{Cov}(Y_1, Y_2 \mid \text{MZ})}{V_P}$$

$$r_{\text{DZ}} = \frac{\mathrm{Cov}(Y_1, Y_2 \mid \text{DZ})}{V_P}$$

Assume that:

- Mating is random (no assortative mating).
- There are no gene-environment interactions or correlations.
- The equal environments assumption holds (MZ and DZ twins experience their shared environments to a similar degree).
- Genetic effects are purely additive (no dominance or epistasis).

2

Assuming the ACE model, demonstrate that the narrow-sense heritability ($h^2 = V_A/V_P$) can be estimated as twice the difference between the MZ and DZ twin correlations.

Show that:

$$h^2 = 2(r_{\mathrm{MZ}} - r_{\mathrm{DZ}})$$

Young, A. I. (2019). Solving the missing heritability problem. *PLOS Genetics* **15**, e1008222.